APPENDIX C

Minutes of the EPA Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) Meeting (October 28 and 29, 2003)

This Page Intentionally Left Blank

SAP Minutes No. 2003-03

October 28 and 29, 2003
FIFRA Scientific Advisory Panel Meeting,
Held at the Holiday Inn Hotel,
Arlington, Virginia

A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding:

Ensuring Data Quality for In Vitro Tests Used as Alternatives to Animal Studies for Regulatory Purposes: A Consultation

Myrta R. Christian, M.S.
Designated Federal Official
FIFRA Scientific Advisory Panel
Date: January 23, 2004

Steven G. Heeringa, Ph.D.
FIFRA SAP, Session Chair
FIFRA Scientific Advisory Panel
Date: January 23, 2004

NOTICE

These meeting minutes have been written as part of the activities of the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP). These meeting minutes represent the views and recommendations of the FIFRA SAP, not the United States Environmental Protection Agency (Agency). The content of these meeting minutes do not represent information approved or disseminated by the Agency. These meeting minutes have not been reviewed for approval by the Agency and, hence, the contents of this report do not necessarily represent the views and policies of the Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use.

The FIFRA SAP is a Federal advisory committee operating in accordance with the Federal Advisory Committee Act and was established under the provisions of FIFRA, as amended by the Food Quality Protection Act FQPA of 1996. The FIFRA SAP provides advice, information, and recommendations to the Agency Administrator on pesticides and pesticide-related issues regarding the impact of regulatory actions on health and the environment. The Panel serves as the primary scientific peer review mechanism of the EPA, Office of Pesticide Programs (OPP) and is structured to provide balanced expert assessment of pesticide and pesticide-related matters facing the Agency. Food Quality Protection Act Science Review Board members serve the FIFRA SAP on an ad hoc basis to assist in reviews conducted by the FIFRA SAP. Further information about FIFRA SAP reports and activities can be obtained from its website at *http://www.epa.gov/scipoly/sap/* or the OPP Docket at (703) 305-5805. Interested persons are invited to contact Larry Dorsey, SAP Executive Secretary, via e-mail at *dorsey.larry@.epa.gov.*

In preparing these meeting minutes, the Panel carefully considered all information provided and presented by the Agency presenters. This document addresses the information provided and presented within the structure of the charge by the Agency.

CONTENTS

October 28 and 29, 2003

Ensuring Data Quality for In Vitro Tests Used as Alternatives to Animal Studies for Regulatory Purposes: A Consultation

PARTICIPANTS

FIFRA SAP, Session Chair
Steven G. Heeringa, Ph.D., Research Scientist & Director for Statistical Design, University of Michigan, Institute for Social Research, Ann Arbor, MI

Designated Federal Official
Mrs. Myrta R. Christian, FIFRA Scientific Advisory Panel Staff, Office of Science Coordination and Policy, EPA

FIFRA Scientific Advisory Panel Members
Stephen M. Roberts, Ph.D. (FIFRA SAP Chair), Professor & Program Director, University of Florida, Center for Environmental & Human Toxicology, Gainesville, FL

Stuart Handwerger, M.D., Director, Division of Endocrinology, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, OH

Gary E. Isom, Ph.D., Professor of Toxicology, School of Pharmacy and Pharmacological Sciences, Purdue University, West Lafayette, IN

Mary Anna Thrall, D.V.M., Professor, Department of Pathology, College of Veterinary & Biomedical Sciences, Colorado State University, Fort Collins, CO

FQPA Science Review Board Members
Joseph P. Bressler, Ph.D., Department of Neurobiology, The Kennedy Krieger Research Institute, Baltimore, MD

George L. DeGeorge, Ph.D., Chief Scientific Officer, MB Research Laboratories, Spinnerstown, PA

James Freeman, Ph.D., ExxonMobil Biomedical Sciences, Inc., Annandale, New Jersey

Michael Luster, Ph.D., Chief, Toxicology and Molecular Biology Branch, National Institute for Occupational Safety and Health, Morgantown, WV

James F. McCormack, Ph.D., Director of Nonclinical Laboratory Compliance, Food & Drug Administration, Rockville, MD

Kenneth Portier, Ph.D., Associate Professor of Statistics, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL

Gregory A. Reed, Ph.D., Associate Professor, Department of Pharmacology, Toxicology, and Therapeutics, University of Kansas Medical Center, Kansas City, KS

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP) has completed its review of the set of scientific issues being considered by the Agency pertaining to the processes for regulatory acceptance of and ensuring the quality of data from *in vitro* tests used as alternatives to animal studies for regulatory purposes. Advance notice of the meeting was published in the *Federal Register* on September 22, 2003. The review was conducted in an open Panel meeting held in Arlington, Virginia, on October 28 and 29, 2003. Dr. Steven G. Heeringa chaired the meeting. Mrs. Myrta R. Christian served as the Designated Federal Official.

The FIFRA SAP was asked to review issues concerned with processes for regulatory acceptance of and ensuring the quality of data from *in vitro* tests used as alternatives to animal studies for regulatory purposes, including performance standards, essential test method components, and quality control of test methods, in the context of three new *in vitro* assays for dermal corrosivity which will be incorporated into its OPTS 870.2500 test guideline for Acute Dermal Irritation.

In preparing these meeting minutes, the Panel carefully considered all information provided and presented by the Agency presenters, as well as information presented by public commenters. These meeting minutes address the information provided and presented at the meeting, especially the response to the charge by the Agency.

CHARGE

Performance Standards

The Agency plans to adopt the Performance Standards developed by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) as a means of communicating the basis by which each of three validated *in vitro* test methods, Corrositex®, EPISKIN™/EpiDerm™, and Transcutaneous Electrical Resistance (TER), are deemed acceptable for providing dermal corrosivity data. Performance Standards consist of descriptions of (1) essential test method components, which are the essential structural, functional, and procedural elements of a validated test method that should be included in the protocol of a proposed mechanistically and functionally similar test method; (2) a minimum list of Reference Chemicals, which is used to assess the accuracy and reliability of the similar test method; and (3) comparable accuracy and reliability values that should be achieved by the proposed test method when evaluated using the minimum set of Reference Chemicals.

Question 1

Please comment on the provisions in the Performance Standards for each of the three methods to demonstrate mechanistic similarity of "me-too" methods. Do the essential test method components for each method adequately describe the unique characteristics of the method necessary to determine whether a test is mechanistically and functionally similar?

Question 2

In its evaluation of any mechanistically similar test system, the Agency plans to use the generic criteria used by ICCVAM for selecting subsets of the Reference Chemicals for all three ICCVAM Performance Standards documents. The criteria specify that chemicals should be selected in such a way that the subset: includes representatives of applicable chemical classes, measures a range of corrosive strengths, includes well-defined chemicals that are currently available commercially, and has unequivocal animal or other *in vivo* evidence. Please comment on the strengths or weaknesses of this approach and identify and discuss any modifications to the criteria that should be considered.

Question 3

The ICCVAM approach for demonstrating functional similarity of "me-too" test methods to validated methods includes the use of well-characterized Reference Chemicals and specifies the accuracy and reliability that should be achieved by "me-too" test systems when tested in intra- and inter-laboratory studies. Please comment on whether "me-too" test systems should be demonstrated to be effective for evaluating the testing endpoint for all of the chemicals in the Performance Standard. Please comment on the value of including chemicals with range of potencies in the Performance Standard. Under what circumstances might testing of "me-too" systems within one laboratory ever be sufficient to demonstrate functional equivalence?

Quality Control

The Agency is proposing quality control measures that should be considered when evaluating the reliability of test kits for regulatory purposes. Please address the following specific issues.

Question 4

Subsets of the Reference Chemicals used in test method validation may be used as training or calibration sets by testing laboratories using *in vitro* systems. Please discuss the utility of and necessity for training or calibration sets in assuring data quality. Please comment on the chemicals selected by ICCVAM for use as a calibration set for TER for this purpose. Please comment on the ranges of chemical classes and potencies of these chemicals. How might other chemicals be selected for possible use in the calibration sets? Please comment on the value of identifying chemicals that might be used by laboratories as training sets to demonstrate proficiency in performing the test.

Question 5

Anticipating the use of systems using tissue constructs, *ex vivo* systems, microarrays or genetically modified cells, please discuss aspects of the quality control criteria that are necessary for assuring the integrity of such systems over time and from lot-to-lot. Please comment on whether and how the type of system - tissue constructs, *ex vivo* systems, or genetically modified cells or animals - should affect the criteria for quality control for assuring the integrity of such systems, both over time and from lot-to-lot.

Question 6

Please comment on the advantages and disadvantages of including concurrent positive and negative controls with *in vitro* assays when used as alternatives to animal testing. What are the important characteristics of positive and negative controls for *in vitro* studies? What aspects of positive control characteristics allow them to be used as part of the quality control process? When might confirmation that positive controls are performing within expected or historical limits be sufficient to demonstrate that the Proprietary Test Method or non-proprietary assay system is functioning properly? When might additional quality control measures be needed?

Question 7

Does the Panel agree that the benchmark controls serve a useful purpose to demonstrate the level of response that can be expected for each chemical class for each lot of Proprietary Test Method assays? Can the Panel suggest criteria for choice of appropriate benchmark controls?

PANEL DELIBERATIONS AND RESPONSE TO CHARGE

The specific issues addressed by the Panel are keyed to the Agency's background documents, and the Agency's charge questions.

Response to Charge

I. Performance Standards

The Agency plans to adopt the Performance Standards developed by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) as a means of communicating the basis by which each of three validated *in vitro* test methods, Corrositex®, EPISKIN™/EpiDerm™, and Transcutaneous Electrical Resistance (TER), are deemed acceptable for providing dermal corrosivity data. Performance Standards consist of descriptions of (1) essential test method components, which are the essential structural, functional, and procedural elements of a validated test method that should be included in the protocol of a proposed mechanistically and functionally similar test method; (2) a minimum list of Reference Chemicals, which is used to assess the accuracy and reliability of the similar test method; and (3) comparable accuracy and reliability values that should be achieved by the proposed test method when evaluated using the minimum set of Reference Chemicals.

Question 1

- Please comment on the provisions in the Performance Standards for each of the three methods to demonstrate mechanistic similarity of "me-too" methods. Do the essential test method components for each method adequately describe the unique characteristics of the method necessary to determine whether a test is mechanistically and functionally similar?

Panel's comments:

The Panel endorsed the Performance Standards (PS) approach to identify and validate "me-too" and "unique" *in vitro* assays. The following paragraphs summarize the Panel's response for each of the three major components of the ICCVAM performance standards for *in vitro* tests.

Structural/functional components:

The Panel concurred that the PS prepared by ICCVAM are very well described for each of the three tests, and the information should provide a basis to determine whether a test is mechanistically and functionally similar to a validated *in vitro* test method. The Panel stated that it would be helpful for the submitting laboratories if the Agency provided examples of what they would consider as a "me-too" assay or a new assay, based upon the essential structural and functional elements (e.g., human skin TER vs. rat skin TER). There was some concern among the Panel members that identification of a "me-too" assay could be a somewhat subjective process rather than one based entirely on objective criteria. However,

with the limited tests that have been evaluated to date (one "me-too" and three unique), there was consensus that this approach of using structural and functional equivalence to determine a "me too" test is conceptually feasible.

Reference Chemicals:

The Panel recommended that NIEHS/EPA (thru ICCVAM) develop a standard list of reference chemicals for validating *in vitro* tests and establish a chemical repository for reference samples/positive controls available to laboratories for developing/conducting *in vitro* skin studies. The reference panel should contain sufficient numbers of different chemical classes (with a range of potency, solubility, etc.) to establish reasonable performance of that specific test for those particular classes of chemicals.

For the three validated test methods, members of the Panel recommended that the laboratories be allowed to determine their own positive control(s) and suggested that the PS not suggest specific examples such as NaOH pellets and 10 N HCl. The Panel felt that these particular examples may be too corrosive, and if suggested by the Agency as a positive control, could become the "gold standard." In lieu of citing specific examples for positive controls, the Panel suggested that the Agency PS provide general requirements (e.g., well characterized, results in a low-to-intermediate response, etc.) wanted in a positive control for a validated test.

For all three validated tests, the Agency PS would benefit from a more thorough discussion of appropriate benchmark controls (range of severity, classes of chemicals) and also how benchmark controls would be considered in the validation studies of the assay. The Panel also recommends that minimum replicate requirements be specified for positive, negative and benchmark controls, and that the PS be unambiguously stated.

Concordance and reliability values:

The Panel suggested that the Agency provide clear guidance on requirements necessary to establish test reliability for the PS for each validated *in vitro* test (how many labs for the inter-laboratory reliability and how many intra-laboratory replications?). The Panel also recommended that the Agency better define what is meant by comparable concordance for test accuracy – will this be statistically based? The Panel expressed the view that the PS should include specific guidelines for minimum achieved sensitivity and specificity of the test when applied to the reference chemical set.

One Panel member expressed the view that if there is no appreciable difference in performance, an *in vitro* assay should be recommended as the preferred alternative testing method for use over an ex-vivo assay (e.g., rat skin TER) as the former more directly addresses the goal of animal replacement.

Question 2

- In its evaluation of any mechanistically similar test system, the Agency plans to use the generic criteria used by ICCVAM for selecting subsets of the Reference Chemicals for all three ICCVAM Performance Standards documents. The criteria specify that chemicals should be selected in such a way that the subset includes representatives of applicable chemical classes, measures a range of corrosive strengths, includes well-defined chemicals that are currently available commercially, and has unequivocal animal or other *in vivo* evidence. Please comment on the strengths or weaknesses of this approach and identify and discuss any modifications to the criteria that should be considered.

Panel's comments:

The Panel expressed the view that the strength of the PS approach to validating a new or "me too" *in vitro* test derives from the stated selection criteria for the Reference Chemical set. By including a range of chemical classes in the Reference Chemical set the general applicability of the test is supported. Choosing Reference Chemicals exhibiting a broad range of corrosive strengths provides insight into the quantitative value of the test. This could be important for assignment of corrosive agents to packing groups. In addition, the inclusion of mildly corrosive agents supports estimation of the sensitivity of the test. The use of well-defined agents with unequivocal animal or other *in vivo* evidence in regard to skin corrosivity anchors the Reference Chemicals as valid "real world" representatives and allows for validated comparisons between the *in vitro* findings and the potential effects of actual environmental or occupational exposures. Limiting the Reference Chemical set to commercially available chemicals allows for the widespread use of this testing regimen.

The Panel identified a weakness of the approach in that it may be difficult to include a sufficient number of Reference Chemicals in each class, both corrosive and non-corrosive, which meet all of these criteria. The Episkin/Epiderm Reference Chemical set comes closest, with 6 of 8 classes containing both corrosive and non-corrosive agents. Although numerous classes of potentially corrosive chemicals are included in the various Reference Chemical sets, some classes are missing. This includes inorganic salts, such as FeCl3, which was reported by ECVAM to be corrosive. Also the Panel noted that hydrocarbons and halogenated hydrocarbons are common solvents and diluents for pesticides, and that these chemicals might be included for study either as individual agents or in combination with other chemicals. The question of how many "classes" the test methods (or "me-too" tests) are validated with, versus the number of classes which the test may be approved for, remains unanswered.

The Panel pointed out that a second weakness of the PS Reference Chemical descriptions for the validated in vitro tests is the lack of standardization of the list. Different groups of specific chemical agents are employed (or recommended) for the different *in vitro* tests. While this may not affect the validation of individual test systems, it does impact on comparisons between the available and proposed test systems.

Question 3

- The ICCVAM approach for demonstrating functional similarity of "me-too" test methods to validated methods includes the use of well-characterized Reference Chemicals and specifies the accuracy and reliability that should be achieved by "me-too" test systems when tested in intra- and inter-laboratory studies. Please comment on whether "me-too" test systems should be demonstrated to be effective for evaluating the testing endpoint for all of the chemicals in the Performance Standard. Please comment on the value of including chemicals with range of potencies in the Performance Standard. Under what circumstances might testing of "me-too" systems within one laboratory ever be sufficient to demonstrate functional equivalence?

Panel's comments:

The Panel agreed that a minimum number of Reference Chemicals (subset of the entire list) should be specified in the PS, to be used for validation procedures of existing alternative test methods, as well as "me-too" tests. It was noted, for example, that there was a large range in the number of Reference Chemicals used among the three test systems presented, with a low of 24 reference chemicals, depending upon the test method under consideration. Although the use of the entire original Reference Chemical set for a validated test method for validation of a "me-too" test might be considered excessive, it is nonetheless important to carry out a sufficiently broad characterization of a new test to validate its performance.

The approach of specifying a known level of accuracy and reliability for a "me-too" test to be considered equivalent to the validated test system was accepted by the Panel. Panel members suggested that Reference Chemicals be limited to those that have been tested with sufficient replication, such that the reliability and accuracy estimates themselves are considered sufficiently precise. The Panel recommended that the concordance of results from "me-too" tests be established by comparison to the unequivocal properties of the test chemicals in human or animal tests, rather than by comparison to an alternative test method. It was recognized that other alternative tests may have less than 100% accuracy (sensitivity and specificity) that would cloud the meaning of "me-too" test "accuracy" or concordance.

One Panel member considered it essential that if as few as 24 or less chemicals are specified in the PS then 100% concordance with *in vivo* test results should be required to demonstrate test equivalence and assure the public safety. Lower percentage concordance would be acceptable if a large enough subset of the Reference Chemicals were tested so as to include more than one chemical from all classes originally validated, with a range of potencies or responses for each class. In the case of the Corrositex validation, a minimal set of 40 reference chemicals were used, resulting in a 25% false positive and 11% false negative rate (Table 2 and Table 3, Section 4.0 and 5.0, respectively, of "ICCVAM Performance Standards: In Vitro Membrane Barrier Test Systems for Skin Corrosion," ICCVAM-DCIWG Proposed MPS; June 23, 2003).

While recognizing that the validated test provides the history (that is, the empirical criteria for acceptable sensitivity, selectivity, etc.), it remains questionable whether this is an

appropriate "bright-line." There may be important statistical or practical considerations to the choice of the subset of Reference Chemicals to be included in the PS. One Panel member queried whether the decision point for qualitative judgment of corrosive agents is sensitive enough to detect even weakly corrosive agents, stating that the judgment of sensitivity cannot be made without validation using known weakly corrosive agents. Thus, the "Performance Standard" should include: 1) a stated minimum number of diverse test chemicals, from all relevant chemical classes; (2) a requirement for Reference Chemicals with varying potencies, efficacies, or range of response, ideally within each chemical class; and 3) minimum standards for reliability and accuracy/concordance in the "me-too" test system results when compared to the known properties of the test chemicals for *in vivo* tests.

A majority of the Panel agreed that validation of a "me-too" test in a single laboratory should be acceptable, if that single laboratory is the only practitioner of the method. The criteria for acceptance should be as rigid as that for a multi-laboratory validation. This would involve at least a sufficient number of independent, repeated tests using the Reference Chemicals to establish the concordance of the "me too" test with a validated test, and to determine the intra-laboratory test reliability of the "me too" test.

The Panel noted the importance of using good experimental design in intra- and inter-laboratory studies, being concerned that there was little discussion of batch-to-batch (or pelt-to-pelt in the case of TER) variability in any of the test method protocols, data, or results. The implication is that this is a very small source of variability for these test systems, which may not be the case in future systems. The general procedures for evaluating "me-too" systems should take this into account.

II. Quality Control

The Agency is proposing quality control measures that should be considered when evaluating the reliability of test kits for regulatory purposes. Please address the following specific issues.

Question 4

- Subsets of the Reference Chemicals used in test method validation may be used as training or calibration sets by testing laboratories using *in vitro* systems. Please discuss the utility of and necessity for training or calibration sets in assuring data quality. Please comment on the chemicals selected by ICCVAM for use as a calibration set for TER for this purpose. Please comment on the ranges of chemical classes and potencies of these chemicals. How might other chemicals be selected for possible use in the calibration sets? Please comment on the value of identifying chemicals that might be used by laboratories as training sets to demonstrate proficiency in performing the test.

Panel's comments:

Given the nature of these *in vitro* systems, particularly in regard to lot-to-lot and day-to-day variability, the Panel felt it essential that test system performance be established and understood. A simple positive and negative control may not be sufficient to represent the

range of responses and the sensitivity required for detection of weakly corrosive agents. In the case of the TER test the twelve Calibration Chemicals suggested by ICCVAM meet the criterion of including strongly and weakly corrosive and non-corrosive agents. However, 12 chemicals constitute a limited test set. It also is incomplete; missing are potentially corrosive inorganic salts like $Fe(Cl)_3$, which is noted in the 60 chemical ECVAM list. Further, the ECVAM "60" list does not completely reflect the classes of chemicals that are important with regard to pesticide registration. Hydrocarbon solvents, for example, find use as diluents but are not included in the list. While most of these solvents are complex mixtures, toxicity profiles can be established both for the mixture and for suitable single-chemical surrogates (e.g., toluene, decane, etc). Clearly a balance must be struck between maintaining a manageable number of Reference Chemicals and assuring that all relevant mechanistic and chemical classes are included.

While the background documents discuss the need for a range of potencies for chemicals, it is important that Reference Chemicals that represent a range of implementation difficulties be included as well. Part of the calibration process for testing laboratories is that the technicians learn to be consistent in application so that reproducible results will be obtained for the Reference Chemicals over time. The potency of a chemical may not be the best measure of how difficult it is for a technician to get consistent results with that chemical. The Reference Chemical set should include some chemicals that are difficult to work with, thereby challenging the technical skill of the staff and forcing them to "stay skilled." Further, some chemicals (e.g., solvents) may destroy the test system; knowledge of this is important if such a chemical is tested in a formulated product.

The Panel noted that training in the use of the validated test is required to be documented under GLPs, presumably with Reference Chemicals. One panel member expressed caution regarding the use of the terms proficiency and calibration set. Proficiency implies a precision and accuracy as may be required by independent accreditation. The training to meet this objective is a laboratory management function. The term, "calibration set" implies traceability to some standard, e.g., a national standard. In the context of the Panel's discussion, Reference Chemicals are identified that can be used as control or benchmark chemicals to help standardize or validate a method in a laboratory and monitor its performance but may not, in the strictest sense, be a true calibration of the test results.

For a training set of chemicals to be used either initially or at some set intervals for the validation of an assay and its performance in a given laboratory, this balance between number of chemicals and inclusivity shifts to a higher number of individual chemicals. Whereas twelve might be an appropriate number for regular "calibration" a training and validation set could easily be 2-3 times this number. This would ensure coverage of relevant classes and potencies for corrosive agents and better test the abilities of a given laboratory to perform the assay accurately.

The Panel expressed the view that Reference Chemical testing:

- Provides relevant training and documentation of training as required by GLPs;
- Provides a means to evaluate technician competency for the test method;

- Permits comparison to a validation database and assessment of variability among labs;
- Identifies relative strengths/weaknesses of the lab and whether additional training is needed.

Question 5

- Anticipating the use of systems using tissue constructs, *ex vivo* systems, microarrays or genetically modified cells, please discuss aspects of the quality control criteria that are necessary for assuring the integrity of such systems over time and from lot-to-lot. Please comment on whether and how the type of system - tissue constructs, *ex vivo* systems, or genetically modified cells or animals - should affect the criteria for quality control for assuring the integrity of such systems, both over time and from lot-to-lot.

Panel's comments:

The use of PS, positive controls, negative controls and benchmark controls will provide the opportunity to achieve a degree of control over the quality of Proprietary Test Methods (PTMs). Two issues that have not been addressed in the PS are how drift in the PTMs will be monitored and how information about problems that arise from the use of these controls will be assimilated and evaluated by the vendor. Individual test facilities may detect failures or out-of-specification performance of the PTM and proceed according to their operating procedures, but the lack of GMP-like regulatory authority does not require these failures to be reported to and addressed by the vendor.

Other facilities may then use an inadequate/under-performing PTM or lot of PTM without benefit of the experiences of the first facility. There should be some consideration that PTM performance reports be compiled by the vendor and reported to purchasers of the PTM. Similar mechanisms are used by computer software vendors to alert purchasers of their products of problems or issues with their products.

The answer to the second part of this question goes beyond the immediate concerns of the Panel, which were *in vitro* tests for corrosive chemicals. Rather, the answer discusses general considerations for future *in vitro* tests that will incorporate the newest advances that are being made in molecular biology. All testing systems require quality control for assuring reproducibility, sensitivity, and specificity. Otherwise, results from the same test repeated in the same laboratory, or in different laboratories, could not be compared. Incorporating positive and negative controls, as well as benchmark samples, monitors quality control. The specific types of controls, the number of controls, the frequency of inserting these controls and the benchmark samples, however, will likely be different for different types of assays. The number of controls would be expected to increase in highly variable systems (e.g. those that require animals) but must be limited because of cost considerations. Hence, the development of newer testing systems that limit variability would have substantial benefit.

An example of the concern for variability is an *ex vivo* system, in which tissue is excised from a donor and cultured as either organ culture, explants, or dissociated cells. There will

be variability in each type of culture because of variability in the donor animals. In primary cultures, however, the variability can be greatly limited if large batches of cells are prepared from several animals and frozen. New testing systems for screening different types of toxic chemicals will likely be developed using genetically modified cell lines. The Ames assay is one example of an already established test that uses genetically modified bacteria to screen for mutagens, which are possible carcinogens. A more complex test system could be developed to establish tissue constructs. For example, a testing system might be developed that uses genetically modified skin stem cell lines that differentiate into skin. The currently available tissue construct uses skin epithelial cells from donors. The advantage of the stem cell line is that the lot-to-lot variability would be reduced because the source of variability, donor tissue, would be reduced.

Microarrays (gene arrays) are powerful endpoint assays that measure changes in the expression of hundreds or thousands of genes and will likely be used in different types of testing systems. One such use would be in classifying xenobiotics according to the patterns of genes that they induce. The pattern of gene expression has been termed a gene fingerprint, and testing systems might be developed for screening xenobiotics by measuring gene fingerprints. In measuring gene fingerprints, rather than one or two specific genes, the testing system has more power for statistical analysis and will likely produce more consistent data. Gene arrays also have the potential of reducing the number of required controls. For example, testing systems for determining gene fingerprints for xenobiotics must use cell lines that express enzymes that metabolize xenobiotics. Positive controls should be incorporated in the testing systems for screening xenobiotics to validate the presence of these enzymes. In using gene arrays, the positive controls might not be necessary because the expression of the activating enzymes, as well as the gene fingerprints, would be determined in the same gene array.

The Panel noted that microarrays and other related systems seem to have a long way to go toward producing reproducible responses among true replicates. In fact, very little true replication is being done, primarily due to the expense of each replicate. As the state of the art in microarray use becomes mature, true replication with demonstrated repeatability may become the standard. When that is truly the case, test systems based on this technology should provide useful tools for risk evaluations. As these new tests are put into practice, more attention must be focused on how drift in performance test standards will be monitored and how information about these problems will be assimilated and evaluated by the vendor.

Question 6

- Please comment on the advantages and disadvantages of including concurrent positive and negative controls with *in vitro* assays when used as alternatives to animal testing. What are the important characteristics of positive and negative controls for *in vitro* studies? What aspects of positive control characteristics allow them to be used as part of the quality control process? When might confirmation that positive controls are performing within expected or historical limits be sufficient to demonstrate that the Proprietary Test Method or non-proprietary assay system is functioning properly? When might additional quality control measures be needed?

Panel's comments:

The Panel commented that insufficient controls may preclude meaningful interpretation of *in vitro* test results. Despite the fact that positive and negative controls are not often used in *in vivo* studies, they are routinely included in *in vitro* studies and it is clearly advantageous and desirable that they be used in the test systems being discussed here. Positive and negative (and vehicle) controls provide needed checks within a study that tell the investigator that the test system appears to be intact and functional. Positive controls help identify performance variability between technicians, between laboratories and between lots of test system. Appropriate controls will likely be needed for some length of time until the Agency and practitioners are satisfied with the performance of the test over time and across laboratories. From a quality control perspective, temporal monitoring of controls across studies and laboratories will help establish consistency of response for the test system.

The Panel again pointed out that a single positive control per assay may not be sufficient and that it may be desirable to include positive control chemicals for one or more of the classification severities. At least one Panel member queried as to what actions should be taken when a negative control produces a positive response or a positive control produces a negative response. The answer will depend on the degree of replication assigned to controls and the specified minimum accuracy or concordance for the test system. Clearly, the developers of the test system should incorporate into the recommended protocols guidance on the degree of replication needed for controls, and what actions should be taken when unexpected results are observed with controls. The degree of replication should be based upon the expected variability and the levels of specificity and sensitivity displayed by the test systems for the Reference Chemicals used as controls. With adequate replication, the fact that positive controls (and negative and vehicle controls, for that matter) are performing within expected limits should be sufficient to demonstrate that the test system is functioning properly.

Question 7

- Does the Panel agree that the benchmark controls serve a useful purpose to demonstrate the level of response that can be expected for each chemical class for each lot of Proprietary Test Method assays? Can the Panel suggest criteria for choice of appropriate benchmark controls?

Panel's comments:

The Panel agrees that benchmark controls are an important mechanism to assess both the adequacy of the method as well as lot-to-lot variability and should be considered as a standard component of these test methods. Benchmark controls, as well as positive and negative controls, should be tested in each new lot to determine the viability and usability of each lot. Control charts could assess variability among lots, and provide a basis for acceptance/rejection. The Panel suggests that benchmark controls include several "classic" responders from different chemical classes/mode of actions.

Variability between different lots is a major concern of the Panel and must be assessed with negative and positive controls as well as benchmark samples. The number of controls and samples depends on several factors; many of which will be defined by the specific test. Certain tests are very consistent and require fewer positive controls and benchmark samples for assessing lot-to-variability whereas other tests are less consistent. The Panel agreed that there is concern regarding whether the lots are large enough to accommodate these types of controls. To address this concern, the Panel suggests that EPA establish the necessary controls and benchmark samples in the individual tests and consults with the manufacturer of the test. Accordingly, the manufacturer would be encouraged to change production so that the size of lots are sufficient for allowing adequate controls and benchmarks.

Appropriate benchmark controls should have the following properties:

- Consistent and reliable source(s) for the chemical
- Structural and functional similarity to the class of article being tested
- Known physical/chemical characteristics
- Supporting data on known effects in animal models
- Known potency in the range of response (including moderate response)

One Panel member stated that benchmark controls can serve a very useful purpose, especially in the situation where the test system demonstrates significant batch-to-batch variability in response. But this variability has not been directly addressed for the test systems being discussed here. If we assume that such variability is quite low, the benefit of re-running benchmark controls for each batch is reduced. In this case, the use of benchmark controls might be relegated to a supplier QC role with periodic running of benchmark chemical to ensure continued consistency of response over time. On the other hand, if the test system does demonstrate significant batch-to-batch variability, it would be important to run benchmark controls more often. Finally, it would seem that benchmark controls would be more important in calibrating a formal dose response model. The need for these controls then depends on the level of precision needed in the final model.

REFERENCES

1. Interagency Coordinating Committee for Validation of Alternative Methods. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised and Alternative Test Methods. NIH Publication 03-4508. September 2003.

2. Interagency Coordinating Committee for Validation of Alternative Methods. 2003. ICCVAM Minimum Performance Standards: In Vitro Membrane Barrier Test Systems for Skin Corrosion.

3. Interagency Coordinating Committee for Validation of Alternative Methods. 2003. ICCVAM Minimum Performance Standards: In Vitro Human Skin Model Systems for Skin Corrosion.

4. Interagency Coordinating Committee for Validation of Alternative Methods. 2003. ICCVAM Minimum Performance Standards: In Vitro Skin Transcutaneous Electrical Resistance (TER) Tests for Skin Corrosion.

5. Interagency Coordinating Committee for Validation of Alternative Methods. 2003. Attachment to memorandum from ICCVAM to OECD dated February 19, 2003. Justification for Development of An OECD Guidance Document on the Application of Good Laboratory Practice (GLP) Principles to *In Vitro* Testing.

6. Interagency Coordinating Committee for Validation of Alternative Methods. 1997. Validation and Regulatory Acceptance of Toxicological Methods of Test Methods. NIH Publication 97-3981. March 1997. See Chapter 2. Validation of Test Methods and Chapter 3. Regulatory Acceptance of Toxicological Test Methods.

7. Interagency Coordinating Committee for Validation of Alternative Methods. 1999. Corrositex®: An *In Vitro* Test Method for Assessing Dermal Corrosivity Potential of Chemicals. NIH Publication No. 99-4495. June 1999.

8. Interagency Coordinating Committee for Validation of Alternative Methods. 2002. ICCVAM Evaluation of EISKIN™, EPIDERM™ (EPI-200), and the Rat Skin Transcutaneous Electrical Resistance (TER) Assay: *In Vitro* Test Methods for Assessing Dermal Corrosivity Potential of Chemicals. NIH Publication No. 02-4502. June 2002. *http://iccvam.niehs.nih.gov/methods/epiddocs/cwgfinal/cwgfinal.pdf.*

9. Hartung, T., Balls, M., Bardouille, C., Blanck, O., Coecke, S., Gstraunthaler, G., and Lewis, D. (2003). Good Cell Culture Practice: ECVAM Good Cell Culture Practice Task Force Report 1. ATLA 30: 407-414.

10. Cooper-Hannan, R., Harbell, J.W., Coecke, S., Balls, M., Bowe, G., Cervinka, M., Clothier, R., Hermann, F., Klahn, L.K., de Lange, J., Lievsch, M. and Vanparys, P. (1999). The principles of Good Laboratory Practice: Application to *in vitro* toxicology studies. The report and recommendations of ECVAM workshop 37. ATLA 27, 539-577.

11. OPPTS Harmonized Test Guideline 870.2600 Acute Dermal Irritation.

12. Environmental Protection Agency. 40 CFR 160 Good Laboratory Practice Standards (Office of Pesticides) July 1, 2003.

13. Environmental Protection Agency. 40 CFR 792 Good Laboratory Practice Standards (Office of Prevention and Toxic Substances). July 1, 2002.

14. Environmental Protection Agency. Documents for FIFRA Scientific Advisory Panel Meeting on Regulatory Use of *In Vitro* Systems, October 28-29, 2003. (This document is a list of documents which appeared on the disk submitted to OSCP on October 2, 2003.

15. Microarray Gene Expression Data Society. Minimum Information About a Microarray Experiment_MIAME for Toxicogenomics (MIAME/TOX). Draft report February 12, 2003.